VU VRIJE UNIVERSITEIT AMSTERDAM

VU Research Portal

(Epi) genetics and twins

van Dongen, J.

2015

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
van Dongen, J. (2015). *(Epi) genetics and twins*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 25. Jan. 2025

Chapter 6

The heritability of DNA methylation in peripheral blood: influences of common SNPs and variability of genetic and environmental variance with age and sex

Abstract

Heritable and environmentally mediated epigenetic variation between individuals may be an important contributor to individual differences in disease susceptibility. We estimated the genetic and environmental variance of DNA methylation level at genome-wide sites in whole blood taking advantage of the resemblance of monozygotic (MZ) twin pairs (N=769) and dizygotic (DZ) twin pairs (N=424) with the classical twin method, and by using estimates of identity-by-descent (IBD) among closely related and distantly related individuals (N=2603, mean age=37.2 years (SD=13.3), 62% females) derived from genome-wide SNPs. The average heritability across genome-wide methylation sites was 19%, of which on average 37% was explained by genome-wide SNPs (MAF > 0.01). Polygenic genotype X sex and genotype X age interaction analysis identified 2654 methylation sites with evidence for sex-specific heritability and 39194 sites with age-specific heritability, and revealed that the between-individual variance due to environmental or stochastic influences increased with age at a large proportion of these sites. The question remains whether these sites affect age and sex-dependent disease susceptibility.

This chapter is based on: Jenny van Dongen [*], Bastiaan T. Heijmans[*], Michel G. Nivard[*], Gonneke Willemsen, Jouke-Jan Hottenga, Quinta Helmer, Conor V. Dolan, Erik A. Ehli, Gareth Davies, BIOS Consortium[iii], H. Eka Suchiman, Rick Jansen, Joyce B. van Meurs, P. Eline Slagboom, Dorret I. Boomsma. The heritability of DNA methylation in peripheral blood: influences of common SNPs and variability of genetic and environmental variance with age and sex. (*manuscript in preparation*)

* These authors contributed equally to the work.

[iii] The Biobank-based Integrative Omics Study (BIOS) Consortium

Introduction

Epigenetic variation between individuals may represent an important contributor to individual differences in disease susceptibility[1]. Of the many epigenetic marks and mechanisms that exist[2], DNA methylation is thought to contribute to stable long-term gene expression regulation and tissue-differentiation[3, 4], and is currently the only that can be assessed at a genome-wide scale in large human epidemiological studies. Epigenome-Wide Association Studies (EWAS) in which a trait, disease or exposure is tested for association with DNA methylation show promising results[5-10]. DNA methylation variation between individuals may result from environmental and stochastic variation or from genetic influences (due to variation in the DNA sequence). Increasing evidence suggests that DNA-sequence-mediated epigenetic variation between individuals contributes to human disease susceptibility[8, 11-13]. Notably, methylation differences have been observed between the sexes[14-16] and across age[17-20], suggesting that epigenetic regulation may also be involved in the widely observed age and sex differences in the aetiology of complex diseases[21, 22]. Studies of prenatal dietary exposures illustrate that the impact of environmental influences on DNA methylation may depend on the sex of the individual[23, 24], while sex-specific SNP effects on DNA methylation have also been described[25].

It is well-known that genetically identical model organisms such as cloned animals[26], isogenic plants[27] and inbred mice[28] exhibit epigenetic and phenotypic differences. Human identical twins provide insight into the extent to which epigenetic regulation in humans may vary due to environmental and stochastic influences. The overall contribution of genetic and environmental differences to variation in DNA methylation between individuals in a population can be estimated by contrasting the correlation between DNA methylation levels of monozygotic (MZ) and dizygotic (DZ) twins, who share 100% and 50% of segregating genetic variants that contribute to methylation differences, respectively (the classical twin design). The average twin-based heritability of DNA methylation across genome-wide CpGs varies between 5% and 19% for different tissues[17, 29-31]. MZ twins already show differences in DNA methylation at birth[31, 32] and some studies have indicated that certain epigenetic marks including DNA methylation at specific loci may diverge in twin pairs with ageing[33, 34], although evidence for such effects is not always seen[35]. Changes in the heritability of DNA methylation with age or between the sexes can be assessed by polygenic genotype X age (or sex) interaction analysis[36].

Here, we report on the largest twin study to date of genome-wide DNA methylation (Illumina 450k array) in whole blood. The study design allows the estimation of the heritability of DNA methylation based on the classical twin method and based on measured genetic relationships between individuals (identity-by-descent, IBD), and to estimate the variance of DNA methylation explained by common SNPs. To obtain insight into potential differences in epigenetic regulation between the sexes and across age, we investigated

polygenic genotype X sex and genotype X age interaction effects on DNA methylation at genome-wide sites. To examine the stability of DNA methylation in blood and the correlation of DNA methylation across two accessible tissues that are suited for human epidemiological studies (blood and buccal), longitudinal and cross-tissue analyses were performed using data from a small subset of subjects.

Materials and Methods

Subjects and samples

The subjects in this study participated in the Netherlands Twin Register (NTR)[37, 38] . In the biobank project, venous blood samples were drawn in the morning after an overnight fast, and separate EDTA tubes were collected for isolation of DNA and assessment of haematological profiles. Blood sampling and buccal sample collection procedures were described in detail previously[39].

Most subjects were twins, but the sample also included parents of twins, siblings of twins or spouses of twins, as described in detail in Appendix 4. In total, 3264 blood samples from 3221 NTR participants were assessed for genome-wide methylation, of which 3089 samples from 3057 subjects passed quality control. Only samples with good quality DNA methylation data and for which data on white blood cell counts were available were kept in the analyses, leaving 3006 samples from 2975 subjects. This dataset included 769 MZ and 424 DZ pairs, and for 31 subjects longitudinal methylation data were available (two time points). All analyses that included genome-wide SNP data were performed on data from a subset of subjects who were genotyped and who were of Dutch origin (N=2603).

For a small subset of 11 MZ pairs (male pairs=3, female pairs=8, age: 18 years), genome-wide methylation data were available for two types of samples: blood (as described above) and buccal. The buccal samples from 10 twins were assessed in 2013, as described by van Dongen *et al*[40]. The 12 additional buccal samples were assessed using the same protocol in 2014. Buccal and blood samples were collected shortly after each other.

All subjects provided written informed consent and study protocols were approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the US Office of Human Research Protections (IRB number IRB-2991 under Federal-wide Assurance-3703; IRB/institute codes, NTR 03-180).

Cell counts

The following subtypes of white blood cells were counted in blood samples: neutrophils, lymphocytes, monocytes, eosinophils, and basophils (see Willemsen et al[39]). Lymphocyte and neutrophil percentages were strongly negatively correlated (r=-0.93). Of these two white blood cell subtypes, the percentage of neutrophils showed the strongest correlation with DNA

methylation levels (as evidenced by the correlation with Principal Components (PCs) from the raw genome-wide methylation data). Basophil percentage showed little variation between subjects, with a large number of subjects having 0% of basophils. Therefore, the percentages of neutrophils, monocytes and eosinophils were used to adjust DNA methylation data for inter-individual variation in white blood cell proportions.

Genome-wide SNP data
Genome-wide SNPs were used to construct a Genetic Relationship Matrix (GRM), which summarizes overall genetic relatedness between all subjects (N=2603), based on all genotyped autosomal SNPs (MAF > 0.01) with the software program Genome-wide Complex Trait Analysis (GCTA)[41]. Details on genome-wide SNP data are provided in the Supplementary Methods.

Infinium HumanMethylation450 BeadChip data
DNA methylation was assessed with the Infinium HumanMethylation450 BeadChip Kit (Illumina, Inc.)[42]. 500ng of genomic DNA from whole blood was bisulfite treated using the Zymo Research 96-well plate using the standard protocol for Illumina 450K micro-arrays, by the department of Molecular Epidemiology from the Leiden University Medical Center (LUMC), The Netherlands. Subsequent steps (i.e sample hybridization, staining, scanning) were performed by the Erasmus Medical Center micro-array facility, Rotterdam, The Netherlands. Quality control and processing of the blood methylation dataset is described in detail in the supplementary methods. In short, a number of sample-level and probe level quality checks were performed. Probes were set to missing in a sample if they had an intensity value of exactly zero, detection *P*-value > 0.01, or bead count < 3. Probes were excluded from all samples if they mapped to multiple locations in the genome[43], had a SNP within the CpG site (at the C or G position) irrespective of minor allele frequency in the Dutch population[44], or if they had a success rate < 0.95 across samples. Only autosomal sites were kept in the analyses. Blood methylation data were normalized with Functional Normalization[45], and normalized intensity values were converted into beta-values (β) and M-values[46]; β -values were used for descriptive purposes only because of their biological interpretability, while M-values were used as input for all analyses. The DNA methylation protocol and data processing of the buccal samples have been described previously [40].

Heritable and environmental influences on DNA methylation levels in blood
To facilitate computations, missing methylation data (0.04%-2.14% of genome-wide probes per individual, mean=0.1%) were imputed with the R package impute. Prior to analysis, the normalized methylation M-values were corrected for sex, age, array row, 96-wells plate (dummy coded), white blood cell percentages (neutrophils, monocytes and eosinophils; assessed at sample collection), and the first ten PCs derived from the genotype data, with the lm

function in R. All analyses that included genome-wide SNP data were performed on the residuals derived after correcting for these covariates. All other analyses (i.e. twin correlations, longitudinal analyses and blood-buccal comparison) were performed on the residuals derived after correcting for the before mentioned covariates minus the genotype PCs.

The proportion of variance in DNA methylation attributable to total additive genetic effects (h^2), the proportion attributable to the additive effects of all measured SNPs (h^2_{SNPs}), and interactions of total additive genetic effects and environmental effects with age and sex were assessed by the classical twin model, and by modeling the DNA methylation data as a function of measured genetic relatedness between subjects. At each CpG site ($CpGi$), the classical twin heritability was computed as:

$h^2(CpGi)$= 2 * (*r*MZ - *r*DZ),

Where *r*MZ and *r*DZ are the correlations of DNA methylation level at one CpG site between the MZ, and between the DZ twins, respectively.

In all other analyses, h^2 was estimated by fitting a linear mixed model in which the covariance of DNA methylation between individuals was modeled as a function of measured genetic relationships based on SNP data. These methods are described in detail by Nivard *et al* (*Nivard M.G., Middeldorp C.M., Lubke G., Hottenga J.J., Abdellaoui A., Boomsma D.I., Dolan C.V. Continuous gene – environment interaction in attention problems, anxious depression, body mass index and height leveraging twin, pedigree and genotype data. manuscript in preparation*). In short, the approach outlined by Zaitlen *et al*[47] was applied, which allows for (simultaneous) estimation of h^2 and h^2_{SNPs} in study samples that include both closely and distantly related individuals. The method makes use of two GRMs: a GRM describing the relationships between all individuals ($\mathrm{GRM}^{\mathrm{IBS}}_{\mathrm{n*n}}$) and a second GRM in which all genetic relationships < 0.05 IBS (distant genetic relationships) are set to zero ($GRM^{IBS>0.05}_{n*n}$), making the estimates of genetic relatedness equivalent to the proportion in the genome shared identity-by-descent (IBD), as explained by Zaitlen *et al*[47]. For each CpG, we jointly estimated the total additive genetic variance (σ^2_{IBD}) and the variance explained by genome-wide SNPs (σ^2_{SNPs}) as follows:

$$var(CpGi) = GRM^{IBS}_{n*n} \otimes \sigma^2_{SNPs} + GRM^{IBS>0.05}_{n*n} \otimes (\sigma^2_{IBD} - \sigma^2_{SNPs}) + I_{n*n} \otimes \sigma^2_e,$$

Where $\mathrm{var(CpGi)}$ is the variance of DNA methylation at CpG i, adjusted for covariates, σ^2_{SNPs} is the variance explained by all SNPs, the term ($\sigma^2_{IBD} - \sigma^2_{SNPs}$) denotes the difference between the total genetic variance and the variance explained by SNPs, and σ^2_e reflects the variance attributable to residual effects (“unique environment”, which may include environmental

influences unique to each individual, stochastic influences and measurement error). The total heritability (h^2_{IBD}) was calculated as: $h^2_{IBD}=\sigma^2_{IBD}/(\sigma^2_{IBD}+\sigma^2_e)$. The proportion of variance explained by genome-wide SNPs was calculated as: $h^2_{SNPs}=\sigma^2_{SNPs}/(\sigma^2_{IBD}+\sigma^2_e)$ and the proportion of the heritability explained by SNPs was calculated as: $h^2_{SNPs}/ h^2_{IBD.}$

Polygenic genotype X sex interaction effects on DNA methylation were investigated with the following model:

$$var(CpGi) = GRM^{IBS>0.05}_{n*n} \otimes \left(\sigma_{IBD} + \beta_{IBD-sex} * Sex\right)^2 + I_{n*n} \otimes \left(\sigma_e + \beta_{e-sex} * Sex\right)^2,$$

Where $\beta_{IBD-sex}$= regression coefficient for the interaction of genetic variance with sex (coded as 0/1), and β_{e-sex}= regression coefficient for the interaction of residual variance with sex (coded as 0/1). This parameterization of the interaction effect is equivalent to how polygenic genotype-by-environment interaction is commonly assessed within the classical twin model, as proposed by Purcell[36].

Polygenic genotype X age interaction effects on DNA methylation were investigated with the following model:

$$\mathrm{var(CpGi)} = \mathrm{GRM}^{\mathrm{IBS>0.05}}_{\mathrm{n*n}} \otimes \left(\sigma_{\mathrm{IBD}} + \beta_{\mathrm{IBD\text{-}Age}} * \mathrm{Age}\right)^2 + \mathrm{I}_{\mathrm{n*n}} \otimes \left(\sigma_{\mathrm{e}} + \beta_{\mathrm{e\text{-}Age}} * \mathrm{Age}\right)^2,$$

Where $\beta_{IBD-Age}$= regression coefficient for the interaction of genetic variance with age (z-transformed), and β_{e-Age}= regression coefficient for the interaction of residual variance with age (z-transformed).

Prior to the analyses based on genome-wide SNP data, methylation data were standardized (z-transformation) to facilitate computations. A small proportion of CpGs for which a model did not run successfully were discarded (see results). The p-values of each of the four interaction effects (genetic and environmental variance by age and sex) were derived with a chi-square test (1 degrees of freedom), where X^2= (beta/se)^2. Statistical significance of interaction p-values was assessed after Bonferroni correction for the number of CpGs for which estimates were successfully obtained. The correspondence between twin-based heritability and heritability estimated on the basis of actual IBD was evaluated by computing the correlation between the value of h^2 for all CpGs based on the classical twin approach and the corresponding value of h^2_{IBD}.

Longitudinal correlation and correlation between blood and buccal methylation

Data from individuals for whom two longitudinal blood samples were collected were used to calculate the correlation between DNA methylation level at time point 1 and DNA methylation level at time point 2 for each CpG site. After

obtaining an estimate of heritability and a longitudinal correlation for each CpG, the correlation between genome-wide estimates of (twin-based) heritability and genome-wide estimates of the longitudinal correlation was estimated to examine the relationship between longitudinal stability and the heritability of DNA methylation level. Data from individuals with 450k methylation data from blood samples and buccal samples were used to calculate the correlation between DNA methylation level in blood and buccal for each CpG. Prior to this analysis, the buccal methylation data (M-values) were corrected for sex, array row and assessment batch (2 levels). Blood-buccal correlations for all CpGs were correlated with the twin-based estimate of h^2 to examine the relationship between the heritability in blood and the extent to which between-individual variation in DNA methylation level is shared across tissues.

Table 1: Characteristics of the subjects.

Sub-group/Analysis	N[A]	Mean age (SD), range	Sex	Interval (years)[C]
MZ twin pairs	769	36.1 (12.4), 18-78	F: 541, M:228	
DZ twin pairs	424	33.9 (10.5), 17-79	FF: 180, MM: 93, FM:151	
Subjects with genome-wide SNP data	2603	37.2 (13.3),17-79	F: 1613, M:889	
Subjects with longitudinal methylation data	31	34.4 (6.1),26-50[B]	F: 24, M:7	5.2 (1.1), 2-7
Subjects with blood and buccal methylation data	22	18[D]	F:16, M:6	

F= Female, M=Male. [A]Number of subjects or twin pairs. [B]Age at first blood sample collection. [C]Time between first and second blood sample. [D]All subjects were 18 years when blood and buccal samples were collected.

Results

Characteristics of the study sample

Genome-wide methylation data were available for 769 MZ twin pairs and 424 DZ twin pairs. The age and sex of subjects is described in Table 1. The analyses with genome-wide SNPs were performed on data from 2603 individuals, including twins (N=2373), parents of twins (N=212), siblings of twins (N=16) and spouses of twins (N=3). Longitudinal data on DNA methylation in peripheral blood samples collected with an interval of on average 5 years were available for 31 individuals and data from blood and buccal samples were available for 22 individuals. Methylation levels at 411169 autosomal CpGs were analyzed. Prior to analyses, DNA methylation levels

were adjusted for age at sample collection, sex, white blood cell percentages, principal components (PCs) from the genotype data, and technical aspects as described in the methods. Together, this set of predictors explained on average 17% of the variance in DNA methylation for genome-wide sites (Supplementary Figure 1).

Table 2 Twin correlations, classical twin heritability, heritability based on IBD, and variance explained by SNPs for DNA methylation level at genome-wide CpGs.

Classical twin approach					Heritability based on GRM				
Parameter	Min	Median	Mean	Max	Parameter	Min	Median	Mean	Max
*r*MZ	-0.14	0.12	0.20	0.99	h_{total}^2	0.00	0.12	0.19	0.99
*r*DZ	-0.25	0.06	0.09	0.89	h_{SNPs}^2	0.00	0.01	0.07	0.98
h^2	-1.56	0.16	0.22	1.65	h_{SNPs}^2 / h_{total}^2	0.00	0.22	0.37	1.00

*r*MZ= Correlation between DNA methylation levels of monozygotic (MZ) twins. *r*DZ= Correlation between DNA methylation levels of dizygotic (DZ) twins. h^2= Heritability of DNA methylation level based on the classical twin method. h_{total}^2= Total IBD-based heritability of DNA methylation level. h_{SNPs}^2= Proportion of variance of DNA methylation explained by genome-wide SNPs (MAF > 0.01). GRM= Genetic Relationship Matrix.

Heritable and environmental influences on DNA methylation

On average across genome-wide CpGs, the pattern of twin correlations (Table 2) suggests that additive genetic influences account for the resemblance of twins for DNA methylation level; the average correlation in MZ twins (*r*=0.20) was approximately twice as large as the average correlation in DZ twins (*r*=0.09). We estimated h^2, the proportion of variance in DNA methylation that can be explained by total additive genetic effects, at each CpG as twice the difference between the MZ and DZ twin correlation. The average heritability estimated by this approach was 0.22. These results are similar to the twin correlations and h^2 based on 450k methylation from peripheral blood published previously[17, 29]. Most CpGs showed little inter-individual variation in DNA methylation (Supplementary Figure 2). The twin correlations and h^2 were on average larger at CpGs with a larger variance (Supplementary Table 1). For example, at the most variable CpGs (defined here as CpGs with a SD of the beta-value >= 0.05, N=28078 CpGs), the average twin-based h^2 was 0.64.

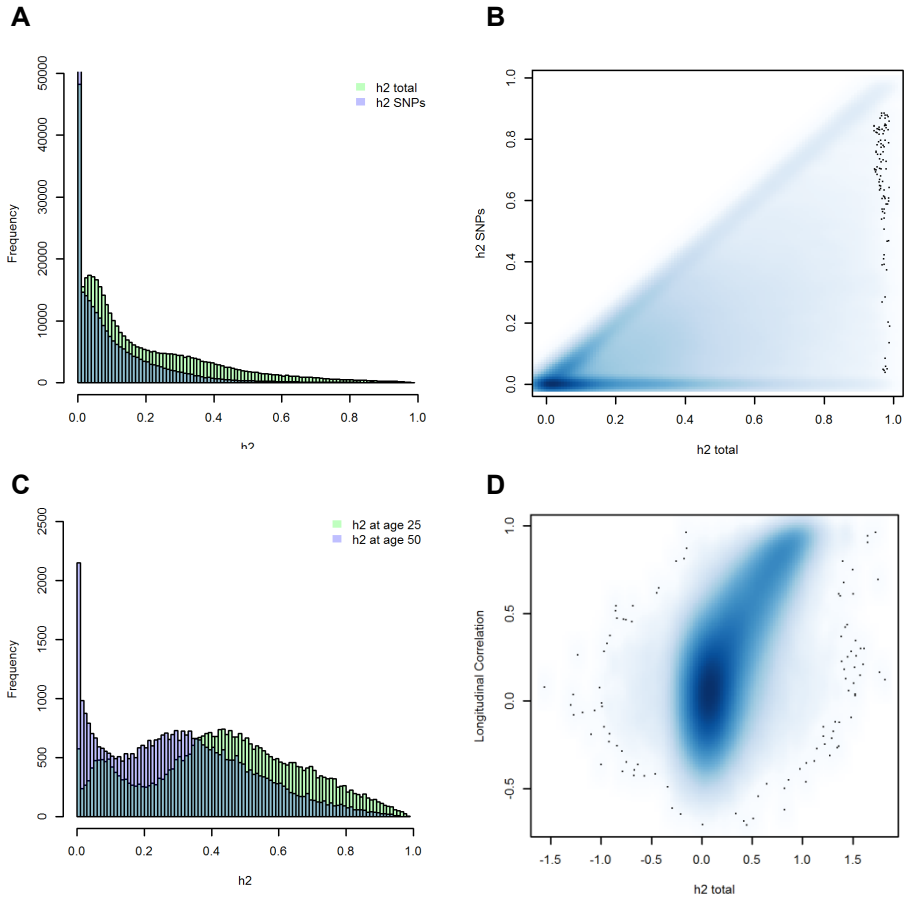
Another approach to estimate the heritability of a trait is to model the trait covariance between subjects as a function of the actual proportion of the genome that they share identity-by-descent (IBD)[48], which can be estimated for example based on genome-wide SNPs. Genome-wide SNP data also allow for estimating the proportion of variation of a trait that can be explained by all

genotyped common SNPS (h^2_{SNPs})[49]. We estimated h^2 and h^2_{SNPs} simultaneously by fitting linear mixed models with two Genetic Relationship matrices (GRMs) to the methylation data of all subjects; one GRM describing genome-wide IBD relationships between closely related (including MZ and DZ twins and their family members) and distantly related individuals (among which IBD is 0), and one GRM describing the genetic relationships between all individuals, respectively[47]. Estimates were successfully obtained for 405010 CpGsites (98.5%). The genome-wide average IBD-based heritability of DNA methylation (h^2=0.19) was similar to the classical twin estimate of h^2, and the estimates from the two methods were strongly correlated (*r*=0.83).

Common SNPs explain a large portion of the heritability of DNA methylation
The average h^2_{SNPs} across all CpGs was 0.07 (Figure 1A), and the proportion of total heritability explained by SNPs (h^2_{SNPs}/ h^2) was on average 0.37 (Table 2). At a relatively large proportion of CpGs, the proportion of total heritability that was explained by SNPs was (almost) 0% or (almost) 100% (CpGs with h^2_{SNPs}/ h^2 < 0.01: 158367 (39%), CpGs with h^2_{SNPs}/ h^2> 0.99=73820 (18%), Figure 1B). These findings indicate that for many CpGs, a relatively large proportion of total estimated heritable variation in DNA methylation can be explained by common SNP effects but there is also a substantial number of CpG sites in the genome where DNA methylation is heritable but cannot be explained by common SNPs. In total, there were 199340 CpG sites (49%) where common SNPs explained at least 1% of the variance of DNA methylation level, and 356741 CpGs (88%) with a total (IBD-based) heritability of at least 1%.

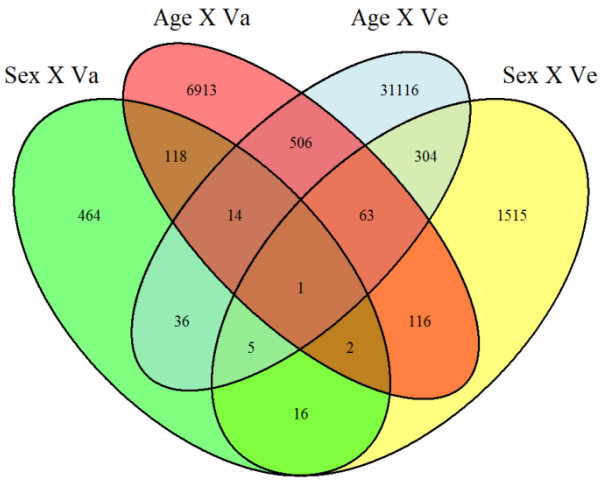
Variation of genetic and environmental influences on DNA methylation by gender and with age
To assess sex differences in the total heritability of DNA methylation, the IBD-based model was extended to allow for interaction between the total genetic variance and sex and interaction between unique environmental variance and sex. Sex interaction models ran successfully for 388950 CpGs (95%). At the genome-wide level, the average heritability was nearly identical in males (mean h^2=0.199, median=0.13) and females (mean h^2=0.198, median=0.13). Significant interaction between sex and the genetic or environmental variance was evident at 2654 CpGs. Of CpGs with significant sex interaction effects, 1562 had a lower heritability in females (59%, Supplemental Figure 3). Compared to all genome-wide sites, CpGs with significant sex interaction had a higher overall heritability on average (mean=0.37, median=0.29) and a larger variance explained by SNPs (mean=0.15, median=0.01), and were overall more variable between subjects (mean SD=0.04 for sites with significant sex interaction, versus mean SD=0.025 for all CpGs). In a similar way, we ran models that included age as a continuous interaction term and found significant interaction of the genetic or environmental variance of DNA methylation with

Figure 1: Heritability of DNA methylation, variance explained by SNPs and longitudinal stability. A): Histogram of total (IBD-based) heritability (h^2total) of DNA methylation (green) and variance explained by genome-wide SNPs (h^2 SNPs, purple) for genome-wide CpGs. The y-axis is truncated. B) Smooth scatterplot of the total (IBD-based) heritability (h^2 total, x-axis) of DNA methylation versus the variance explained by genome-wide SNPs (h^2 SNPs) for genome-wide CpGs. The density of data points is denoted by the color intensity (darker blue=higher density of data points). C) Histogram of IBD-based heritability at age 25 (green) and at age 50 (purple) for 39194 CpGs with significant interaction between age and genetic variance or between age and unique environmental variance. Dark blue denotes the overlap of green and purple bars. D) Smooth scatterplot of (twin-based) total heritability (h^2 total, x-axis) versus longitudinal correlation (y-axis) for genome-wide CpGs.



age at 39194 CpGs. Age interaction models ran successfully for 375783 CpGs (91%). CpGs with significant age interaction were more variable on average between subjects (mean SD age moderated CpGs= 0.036), had a higher total heritability on average (mean=0.37, median=0.36, based on IBD; mean=0.47, median=0.51, based on twins), and a larger proportion of variance explained by SNPs (mean=0.11, median=0.05). To illustrate the direction of effect, heritability of DNA methylation at age 25 and 50 is plotted for sites with significant age interaction effects in Figure 1C. For 90% of sites with significant age interaction, the heritability was lower at age 50 than at age 25. There were 2147 CpGs where the heritability was < 0.01 at age 50, suggesting that genetic variation contributes very little to the variation in methylation at these sites at a higher age, while the heritability of methylation at these sites at age 25 was on average 0.21. For most sites, the change in heritability was modest (Supplementary Figure 4), but large differences also occurred. For example, there were 103 CpGs where the change in heritability was larger than 0.5 between age 25 and age 50. For 32045 sites (82 %) with significant age moderation and 2022 sites (76%) with significant sex moderation, the difference in heritability (which is expressed as the ratio of additive genetic effects over the sum of additive genetic effects plus unique environmental effects) was related to a change in the unique environmental component. The overlap of age and sex moderation effects for individual CpGs is illustrated in Figure 2.

Figure 2: Venn diagram of the number of CpGs with significant interaction between sex and genetic variance (Sex X Va), between sex and unique environmental variance (Sex X Ve), between age and genetic variance (Age X Va) and between age and unique environmental variance (Age X Ve).



Longitudinal correlation of DNA methylation in blood

The longitudinal correlation was on average 0.21 (median=0.16), and the longitudinal correlation was strongly correlated with the h^2 of methylation at the same CpG (*r*=0.70), which suggests that sites with a larger heritability tend to be more stable across the time interval studied, as previously observed[50]. Highly heritable methylation sites ($h^2 > 0.8$, N= 17871) were highly stable over time (mean *r*=0.79, median *r*=0.84). Of note, sites with a high longitudinal correlation and a low h^2 also exist (Figure 1D): at these sites, longitudinal stability may reflect stability of environmentally driven variation or stability of variation that initially arose stochastically.

DNA methylation correlation between blood and buccal cells

Of 405487 CpGs common to the blood and buccal dataset, methylation level was positively correlated between blood and buccal at 244703 CpGs (mean *r*=0.22, median *r*=0.19) and negatively correlated at 160784 CpGs (mean *r*=-0.16, median *r*=0.14) between blood and buccal samples. Across all CpGs, the heritability in blood was weakly correlated with the size of the cross-tissue correlation (*r*=0.32). Focusing only on the most variable methylation sites in blood (SD of the β-value >= 0.05), heritability in blood was moderately correlated (*r*=0.54) with the blood-buccal correlation. Of note, CpGs where methylation level was strongly correlated between blood and buccal ($r > 0.8$, N=1015 CpGs, or $r < -0.8$, N=2 CpGs) were highly heritable and highly stable over time in blood (mean h^2=0.91, median=0.93; mean longitudinal *r*=0.83, median *r* =0.90). These findings suggest that when inter-individual variation in methylation level is strongly correlated between blood and buccal cells, genetic influences are main the cause of this correlation.

Discussion

We assessed genome-wide DNA methylation in whole blood in a large population-based twin cohort (including a small group of family members of twins) and estimated the heritability of DNA methylation level based on the classical twin method and based on measured genetic relationships between individuals (IBD). The two methods provided similar estimates: the genome-wide average heritability was 0.22 based on the classical twin approach and 0.19 based on IBD. Across genome-wide CpGs, the average proportion of variance of DNA methylation level explained by common SNPs was 0.07, and SNPs explained on average 37 % of the total heritability of methylation level. Our results indicate that for 49% of CpGs targeted by the 450k array at least 1% of the variation of DNA methylation in whole blood between individuals can be ascribed to the effects of common SNPs. These findings provide guidance for methylation QTL analyses by giving an indication of the number of CpGs that may be identified through methylation QTL analyses on whole blood samples (given sufficient power). Interestingly, at 18% of 450k targeted CpGs, > 99% of the heritability was explained by common SNPs. Yet, at the genome-

wide level there was a discrepancy between total heritability and variance explained by common SNPs. The proportion of DNA methylation heritability that cannot be explained by common SNPs may derive from genetic variants that are not or incompletely tagged by common SNPs, including rare variants and structural variation. CpGs with a higher heritability of DNA methylation tended to show larger stability across a time interval of on average 5 years. Also, CpGs where between-individual variation in methylation level was strongly correlated between blood and buccal cells were characterized by a high heritability in blood, suggesting that genetic influences underlie strong cross-tissue correlations.

We identified 2654 sites with evidence for sex-specific heritability and 39194 methylation sites with age-specific heritability. In support of previous indications that certain epigenetic marks between monozygotic twins may diverge as the twins age[33, 34] (a phenomenon referred to as epigenetic drift), our study indicated that the majority of methylation sites with significant age interaction showed a decreasing heritability with age due to increasing importance of environmental or stochastic influences on DNA methylation. The question remains whether these sites affect age and sex-dependent disease susceptibility.

Reference List

1. Mill,J. & Heijmans,B.T. From promises to practical strategies in epigenetic epidemiology. *Nat. Rev. Genet.* **14**, 585-594 (2013).
2. Bird,A. Perceptions of epigenetics. *Nature* **447**, 396-398 (2007).
3. Cedar,H. & Bergman,Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10**, 295-304 (2009).
4. Meissner,A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-770 (2008).
5. Dick,K.J. *et al.* DNA methylation and body-mass index: a genome-wide analysis. *Lancet* **383**, 1990-1998 (2014).
6. Hidalgo,B. *et al.* Epigenome-wide association study of fasting measures of glucose, insulin, and HOMA-IR in the Genetics of Lipid Lowering Drugs and Diet Network study. *Diabetes* **63**, 801-807 (2014).
7. Irvin,M.R. *et al.* Epigenome-wide association study of fasting blood lipids in the Genetics of Lipid-lowering Drugs and Diet Network study. *Circulation* **130**, 565-572 (2014).
8. Liu,Y. *et al.* Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* **31**, 142-147 (2013).
9. Suter,M. *et al.* Maternal tobacco use modestly alters correlated epigenome-wide placental DNA methylation and gene expression. *Epigenetics.* **6**, 1284-1294 (2011).
10. Zeilinger,S. *et al.* Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS. One.* **8**, e63812 (2013).
11. Gamazon,E.R. *et al.* Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Mol. Psychiatry* **18**, 340-346 (2013).

12. Zhang,X. *et al.* Linking the genetic architecture of cytosine modifications with human complex traits. *Hum. Mol. Genet.* **23**, 5893-5905 (2014).
13. Shi,J. *et al.* Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. *Nat. Commun.* **5**, 3365 (2014).
14. Boks,M.P. *et al.* The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS. One.* **4**, e6767 (2009).
15. Tapp,H.S. *et al.* Nutritional factors and gender influence age-related DNA methylation in the human rectal mucosa. *Aging Cell* **12**, 148-155 (2013).
16. Liu,J., Morgan,M., Hutchison,K., & Calhoun,V.D. A study of the influence of sex on genome wide methylation. *PLoS. One.* **5**, e10028 (2010).
17. Bell,J.T. *et al.* Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS. Genet.* **8**, e1002629 (2012).
18. Horvath,S. *et al.* Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol.* **13**, R97 (2012).
19. Horvath,S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
20. Teschendorff,A.E. *et al.* Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res.* **20**, 440-446 (2010).
21. Kaminsky,Z., Wang,S.C., & Petronis,A. Complex disease, gender and epigenetics. *Ann. Med.* **38**, 530-544 (2006).
22. Bjornsson,H.T., Fallin,M.D., & Feinberg,A.P. An integrated epigenetic and genetic approach to common human disease. *Trends Genet.* **20**, 350-358 (2004).
23. Lillycrop,K.A., Phillips,E.S., Jackson,A.A., Hanson,M.A., & Burdge,G.C. Dietary protein restriction of pregnant rats induces and folic acid supplementation prevents epigenetic modification of hepatic gene expression in the offspring. *J. Nutr.* **135**, 1382-1386 (2005).
24. Tobi,E.W. *et al.* DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum. Mol. Genet.* **18**, 4046-4053 (2009).
25. Flores,K.G. *et al.* Sex-specific association of sequence variants in CBS and MTRR with risk for promoter hypermethylation in the lung epithelium of smokers. *Carcinogenesis* **33**, 1542-1547 (2012).
26. Rideout,W.M., III, Eggan,K., & Jaenisch,R. Nuclear cloning and epigenetic reprogramming of the genome. *Science* **293**, 1093-1098 (2001).
27. Cortijo,S. *et al.* Mapping the epigenetic basis of complex traits. *Science* **343**, 1145-1148 (2014).
28. Morgan,H.D., Sutherland,H.G., Martin,D.I., & Whitelaw,E. Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* **23**, 314-318 (1999).
29. McRae,A.F. *et al.* Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol.* **15**, R73 (2014).
30. Grundberg,E. *et al.* Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. *Am. J. Hum. Genet.* **93**, 876-890 (2013).
31. Gordon,L. *et al.* Neonatal DNA methylation profile in human twins is specified by a complex interplay between intrauterine environmental and genetic factors, subject to tissue-specific influence. *Genome Res.* **22**, 1395-1406 (2012).

32. Martino,D. *et al.* Longitudinal, genome-scale analysis of DNA methylation in twins from birth to 18 months of age reveals rapid epigenetic change in early life and pair-specific effects of discordance. *Genome Biol.* **14**, R42 (2013).
33. Fraga,M.F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *PNAS* **102**, 10604-10609 (2005).
34. Talens,R.P. *et al.* Epigenetic variation during the adult lifespan: cross-sectional and longitudinal data on monozygotic twin pairs. *Aging Cell* **11**, 694-703 (2012).
35. Bocklandt,S. *et al.* Epigenetic predictor of age. *PLoS. One.* **6**, e14821 (2011).
36. Purcell,S. Variance components models for gene-environment interaction in twin analysis. *Twin. Res.* **5**, 554-571 (2002).
37. Boomsma,D.I. *et al.* Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* **9**, 849-857 (2006).
38. Willemsen,G. *et al.* The adult Netherlands twin register: twenty-five years of survey and biological data collection. *Twin. Res. Hum. Genet.* **16**, 271-281 (2013).
39. Willemsen,G. *et al.* The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. *Twin Res Hum Genet* **13**, 231-245 (2010).
40. van Dongen,J. *et al.* Epigenetic variation in monozygotic twins: a genome-wide analysis of DNA methylation in buccal cells. *Genes (Basel)* **5**, 347-365 (2014).
41. Yang,J., Lee,S.H., Goddard,M.E., & Visscher,P.M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76-82 (2011).
42. Bibikova,M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98**, 288-295 (2011).
43. Chen,Y.A. *et al.* Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics.* **8**, 203-209 (2013).
44. Genome of the Netherlands Consortium Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature Genetics* **46**, 818-825 (2014).
45. Fortin,J.P. *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *bioRxiv*(2014).
46. Du,P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC. Bioinformatics.* **11**, 587 (2010).
47. Zaitlen,N. *et al.* Using extended genealogy to estimate components of heritability for 23 quantitative and dichotomous traits. *PLoS. Genet.* **9**, e1003520 (2013).
48. Visscher,P.M. *et al.* Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS. Genet.* **2**, e41 (2006).
49. Yang,J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**, 519-525 (2011).
50. Shah,S. *et al.* Genetic and environmental exposures constrain epigenetic drift over the human life course. *Genome Res.* **24**, 1725-1733 (2014).